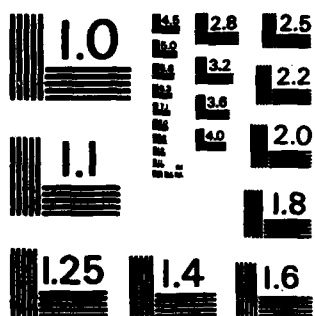AD-A162 585

ACCURACY IN PERFORMANCE APPRAISALS: A COMPARISON OF TWO
RATER COGNITIVE PROCESS MODELS(U) TEXAS A AND M UNIV
COLLEGE STATION S L FRANK-MAJOR ET AL. OCT 85 TR-ONR-1

1/1

UNCLASSIFIED N00014-85-K-0289

F/G 5/9

NL

END

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD-A162 585

⑫

Organizational Behavior Research

Department of Management

Department of Psychology

ACCURACY IN PERFORMANCE APPRAISALS:
A COMPARISON OF TWO RATER
COGNITIVE PROCESS MODELS

Susan Lee Frank-Major
and
Roseanne J. Foti

October 1985
TR-ONR-1

DTIC
ELECTE
DEC 20 1985
S   D
B

DTIC FILE COPY

Texas A&M University

85 12 20 007

ACCURACY IN PERFORMANCE APPRAISALS:
A COMPARISON OF TWO RATER
COGNITIVE PROCESS MODELS

Susan Lee Frank-Major
and
Roseanne J. Foti

October 1985
TR-ONR-1

Department of Psychology
Department of Management
Texas A&M University

DTIC
ELECTE
DEC 20 1985
B

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR-ONR-1	2. GOVT ACCESSION NO. AD-A162 585	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ACCURACY IN PERFORMANCE APPRAISALS: A COMPARISON OF TWO RATER COGNITIVE PROCESS MODELS.		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Susan Lee Frank-Major and Roseanne J. Foti		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0289
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Texas A&M University College Station, Texas 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 475-036
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA 22217		12. REPORT DATE October, 1985
		13. NUMBER OF PAGES 54
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Supported by the Office of Naval Research Manpower R & D Program

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Performance appraisal, rating accuracy, cognitive categorization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two different theories of the cognitive processes involved in rating
performance were compared by Nathan and Lord in 1983. These theories
comprised Borman's (1978) traditional model of dimensional schemata
and Feldman's (1981) cognitive categorization theory. To further explore
the role of each in the process of performance appraisal over time, par-
ticipants in the present study were presented with two different video-
tapes of a lecturing college instructor. One half of the participants

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-LF-014-6601

initially viewed a tape exhibiting primarily good lecture behaviors and two days later viewed a videotape containing mostly poor performance behaviors. The order of videotape presentation was manipulated so that the remaining participants viewed the tapes in reverse order. Subsequently, subjects appraised the lecturer's performance on a series of Likert rating scales and also completed a questionnaire examining their ability to recall specific performance behaviors. These measures revealed partial support for both theories. In general, the performance ratings lent support to the traditional model while cognitive categorization was corroborated in the recognition memory task.

QUAL...
INSPECTED
3

Accession For		
NTIS GRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
J.		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1		

Performance Appraisal

Cascio (1982) has defined performance appraisal as "the systematic description of individual job-relevant strengths and weaknesses" (p. 309). Appraisals are conducted in organizations for numerous purposes. For example, they are frequently used as a basis for promotion and placement, as a criterion to validate selection devices and training programs, or as a basis for rewards and feedback (Kane & Lawler, 1979). The performance appraisal is viewed as a function of three interacting systems: the organizational setting within which the appraisal occurs, the appraiser's capabilities to process information, and the appraisee's behavioral patterns (Ilgen & Feldman, 1983). Facets of each system can contribute to inaccurate and biased evaluations. The appraisal of employee performance can directly affect not only the individual being evaluated, but also the maintenance of the organization's effectiveness (Latham & Wexley, 1980). For these and other reasons, it is critical that performance evaluations be as accurate as possible. Unfortunately, this procedure has been plagued with many deficiencies which preclude flawless appraisals.

In general, there are two types of performance measures--the objective (nonjudgmental) and the subjective (judgmental). Nonjudgmental data include measures of production output, errors, and task completion times, as well as records of absenteeism, turnover, grievances, and accidents (Landy & Farr, 1983). Performance in the majority of jobs, however, is not easily measured in objective terms. Generally, reliance on nonjudgmental measures will not adequately

capture the essence of employees' performance. Judgmental data, on the other hand, allow for a wider range of discretion and application. Many researchers have found that subjective measures, specifically rating scales, are used by an overwhelming majority of organizations (e.g., Bigoness, 1976; Borman, 1979; DeNisi & Stevens, 1981). Indeed, ratings are the most ubiquitous form of performance appraisal.

Performance Ratings

One cannot assume from the wide use of ratings that they are the method least susceptible to error. Rather, they are inevitably contaminated by a host of problems.

In a comprehensive review of the literature on performance rating, Landy and Farr (1980) asserted that three general variables influence the rating process: (a) the roles of the rater and ratee, (b) the vehicle or rating instrument, and (c) the rating context. Investigations into each of these factors have differentially contributed to our understanding of the appraisal process.

The vehicle, or rating format, has received much attention in research and literature. The basic assumption underlying these studies is that the type of format chosen may affect the accuracy and adequacy of evaluations. As several authors have suggested, the conclusions reached have been less than enlightening in finding the superior vehicle (e.g., Bernardin & Cardy, 1982; DeNisi & Stevens, 1981).

For example, in 1979, Borman studied the effects of rating format (and rater training) on accuracy in performance ratings. Selecting

five formats he believed to be most promising, student subjects were asked to evaluate the job performance effectiveness of a manager or recruiter. The results of a Job X Format interaction indicated no superiority of one format over any other.

The bleak conclusion reached by many researchers has been that "after more than 30 years of serious research, it seems that little progress has been made in developing an efficient and psychometrically sound alternative to the traditional graphic rating scale" (Landy & Farr, 1980, p. 89). It appears that the rating format, itself, is less important than the cognitive effect it has on the rater.

As if this conclusion is not disheartening enough, research into the second component of the rating process (namely, the rating context) has not fared much better. However, unlike the rating format, one of the central problems associated with rating context is a relative lack of research. Included in this area are studies investigating the effects of the intended use of performance evaluations as well as position characteristics. At least one major conclusion can be drawn. It appears that ratings for administrative purposes tend to be more lenient and less accurate than those obtained for research purposes or employee development (DeCotiis & Petit, 1978; Warmke & Billings, 1979; Zedeck & Cascio, 1982). Unfortunately, most investigations in this area have been conducted for research purposes and therefore do not provide decisive conclusions about the impact of rating purposes within the organization. Rating variances do not appear to be highly dependent on the purpose component, however, more efficacious tests are needed (Landy & Farr, 1980).

Lest the reader give up hope on the rating process, attention will now be turned to the third component--roles--which has enjoyed much attention and greater success in reaching beneficial conclusions. This variable can be divided into three constituents: ratee characteristics, rater characteristics, and an interaction of the two. A comprehensive review of these are beyond the scope of this paper, however, a brief overview follows. For further information, readers are directed to the literature by Cascio (1982), Dunnette and Borman (1979), and Landy and Farr (1983).

Ratee characteristics include personal factors such as race, age, and sex; and job-related variables such as performance level, tenure, and reaction to performance appraisal. Rater characteristics again include race, age, and sex, but also embrace intellectual skills, job experience, knowledge of the rater and the job, as well as numerous other factors.

To many laypersons it may appear that ratings should be more a function of the ratee's real performance than of the characteristics of the one doing the rating. However, as Bernardin and Cardy (1982) have stated, "there is a strong indication that ratings are as much or more a function of the idiosyncracies of the rater who made them than they are of the actual behavior of the ratees" (p. 352).

Of particular interest to the present project is the cognitive characteristics of the rater. Recently, much attention has been devoted to this area, and rightfully so. The cognitive characteristics of raters seem to be the key to significantly increasing our understanding of the rating process. Both industrial and social psy-

chologists have explored the mental processes underlying our evaluations of others.

This current surge of interest in a cognitive approach to performance apparisal is evidenced in a model proposed recently by DeNisi, Cafferty, and Meglino (1984). Their model consists of a series of interrelated steps reflecting the notion that the appraisal process is a judgmental activity dependent on social perception and cognition. This model has a unique perspective in that it views the rater as actively seeking the information required to formally evaluate performance. Accordingly, the steps involved in appraisal include: (a) observing employee behavior, (b) cognitively representing that behavior, (c) storing the representation in memory, (d) retrieving the information required to make a formal evaluation, (e) examining and integrating additional pieces of information with the retrieved data, and (f) formally appraising the employee with the use of a rating instrument.

In addition to presenting their model, DeNisi et al. (1984) suggest several research propositions. As a part of this, they call for an investigation to determine whether raters are more likely to recall specific ratee behaviors or only overall impressions when appraising performance. There are theories which address this issue either indirectly or directly. This proposal focuses on two such theories of cognitive processing which have recently been compared by Nathan and Lord (1983). It is believed that through determination of which model more appropriately fits the rater's cognitive process, accurate ratings will be a goal less distant.

Cognitive Categorization Model

Feldman (1981) developed a categorization model in order to explain the appraisal process, which he viewed as a more specific case of general cognitive processes. This model is therefore steeped in cognitive-social psychology. It is also an outcome of research conducted by Rosch and her coworkers (Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) and related explorations by Cantor and Mischel (1977, 1979).

Schemas. To oversimplify, the cognitive categorization model asserts that people classify and organize information about others into general categories. This allows us to integrate, discriminate, and simplify an enormous amount of information by attributing stability to another's behavior (Snyder & Uranowitz, 1978). In short, it makes life easier to contend with a neat little package of who "Mrs. Smith" is rather than sift through and store all the different pieces of information about her.

> A schema is a cognitive structure that consists in part of the representation of some defined stimulus domain, including a specification of the relationships among its attributes, as well as specific examples or instances of the stimulus domain. As such, one of the chief functions of a schema is to provide an answer to the question, "what is it?" (Taylor & Crocker, 1981, p. 91).

A schema, therefore, allows us to compartmentalize people, and if asked about them we can activate the category we have placed them in rather than sort through all of our specific interactions with them.

Furthermore, a category prototype is developed against which judgments of similarity can be made. The prototype is an abstraction of the most representative and inherent features of a schema. For example, if you know that Mrs. Smith can prepare great pies, personally maintains a tidy home, has raised three children, and needle points her sofa pillows, you may categorize Mrs. Smith into your "traditional homemaker" schema because she exhibits behaviors that are prototypically associated with this category. Furthermore, upon remembering Mrs. Smith, you may also attribute to her an ability to cook pot roast, mend clothes, and tell bedtime stories, not because you have observed her doing them, but rather because they are additional behaviors categorized under your "homemaker" schema.

Judgmental applications. Once an evaluation has been made it is subsequently used as a basis for later inferences about the person (Srull & Wyer, Jr., 1979) and, as many researchers have found, it is extremely difficult to change these initial judgments (e.g., Cantor & Mischel, 1979; Lord, Ross, & Lepper, 1979). In fact, commitment to our incipient categorization of others is so prevalent and powerful that it may color any future inconsistencies in their behavior. Once an employee is categorized, further judgments of that employee are influenced by the category prototype. This process may produce under-evaluations and/or overevaluations of the employee (Feldman, 1981).

Ostrom, Lingle, Pryor, and Geva (1980) discuss a number of their studies concerned with the cognitive categorization of person impressions. One experiment (Lingle & Ostrom, 1979) focused on the influence an initial judgment has on the ease of making a subsequent judgment.

Participants were asked to judge the appropriateness of a stimulus person for a designated vocation. Decisions were based on a set of stimulus traits describing the target person and were presented during this decision-making process. The participants were then asked to make another decision regarding the suitability of the target person for an occupation requiring similar or dissimilar traits to that of the first occupation. Furthermore, the descriptive traits were removed during this decision-making process so that the judgment was based on memory. The speed with which subjects made this second judgment served as the dependent variable. Results indicated that subjects required more than a second longer to make the occupational judgment involving dissimilar traits as opposed to those requiring similar traits. This finding suggests that an initial thematic decision has an influential impact on subsequent judgments.

In a related study, Lingle (1979) asked subjects to judge whether a specified trait would characterize a target person that had been already characterized by two other traits. Of these two traits, one trait was relevant and the other irrelevant to making the judgment. After making their decision, participants performed a distractor task for 50 seconds. They were then asked to make an occupational judgment based on memory of the stimulus person. While contemplating their decision, subjects were interrupted by a probe word that was nearly illegible. This word was either one of the three descriptive trait words used in the earlier stages of the experiment or a control trait that had not been associated with the target person. A faster recognition speed was established for probe words previously associated with

the target as opposed to unassociated probes. Furthermore, the recognition times revealed that subjects accessed relevant stimulus traits more readily than irrelevant traits during the judgment interval.

Based on these results, the authors concluded that initial judgments guide and determine later evaluation. "Most striking is the persistent evidence that an initial judgment, rather than factual stimulus information, is remembered and used as the basis for subsequent judgments" (Ostrom, Lingle, Pryor, & Geva, 1980, p. 84). This suggests that people rely on earlier categorization, instead of specific pieces of information when appraising others. In effect, categorization not only colors and biases recall, it also precludes contradictory evidence from surfacing and, in fact, elicits confirming evidence. Thus, testing an impression necessitates a search for supporting evidence and, consequently makes disconfirmation of the impression more improbable (Feldman, 1981).

Kulik (1983) further supports the argument that initial beliefs about others persevere even in the face of contradictory behavior. In Kulik's study, subjects first watched a videotape of two people getting acquainted, one of which was a target person. During this initial exposure, subjects developed either extroverted or introverted schematic impressions of the target. Next, a second videotape was viewed which confirmed or disconfirmed the initial schematic impression. Even when disconfirming behavior was viewed, initial images were maintained. Kulik (1983) concludes that our beliefs about others are not likely to change "as a simple function of impartially tallying each instance of consistent and inconsistent behavior" (p. 1978);

instead, persistence of schematic impressions seems to be the rule rather than the exception.

Foti, Fraser, and Lord (1982) have demonstrated the significance of categorization in perceptions of political leaders. The basic assumption underlying their research was that the perception of a leader is developed by comparing the person being judged with leader prototypes. Once categorized, further evaluations will be founded on the category prototype rather than actual behaviors. In part, Foti et al. (1982) concluded that prototypes operate much like stereotypes in that they specify characteristics associated with category members.

Halo. One by-product of cognitive categorization is a phenomenon designated by Thorndike (1920) as the halo effect. In the rating process, halo refers to a tendency to attend to global impressions of the ratee rather than to differentiating levels of job behavior.

Cooper (1981) distinguishes between two types of halo; true and illusory. True halo refers to the degree of co-occurrence actually arising in ratees' skills or covariance between rating categories (Fox, Bizman, & Herrmann, 1983). Illusory halo exists when observed halo surpasses true halo. The rater perceives a degree of covariance or co-occurrence not actually reflected among ratees or the dimensions. Investigation interests focus on the latter type of halo in lieu of the former.

Schemas can, and do incorporate both types of halo. For example, our homemaker schema may contain an element delineating the female homemaker as a good seamstress and also contain an element depicting

the homemaker as a charming social hostess. However, any one specific woman that is categorized under the homemaker schema, may or may not possess these qualities. If in fact she is socially inept but we are so dazzled by the radiance of her other homemaker attributes we will tend to cast a halo around all schema-related qualities. Therefore, a part of our perception is based on "true" halo in that she may actually be an excellent seamstress and a good cook. However, our perception is also influenced by "illusory" halo such that we attribute to her an ability to entertain when, in reality, she does not possess this ability.

Halo, in effect, is an outgrowth of schema processing. There is a tendency to overestimate the information that is consistent with a particular schema and simultaneously underestimate the evidence that is inconsistent or irrelevant (Taylor & Crocker, 1981).

Halo can be particularly problematic for performance ratings in that raters may rely on their schemas to determine the ways in which rating categories covary. Furthermore, unless disconfirming evidence is salient and acted on, ratings will covary among putatively related categories (Cooper, 1981).

As part of a longitudinal study spanning five successive rating periods and three and one-half years, the source and stability of halo was examined by Vance, Winne, and Wright (1983). Data were collected in a metropolitan police department as part of an ongoing performance appraisal program. Results indicated that reliable halo variance stemmed from the behavior of raters rather than from ratees. The authors report that this finding underscores the importance of

the rater in the performance appraisal process.

Much research has been done to try to eliminate or reduce the effects of halo in ratings (e.g., Bernardin & Pence, 1980; Cooper, 1983; Johnson, 1963; Johnson & Vidulich, 1956; Kenny & Berman, 1980). Some research findings have been more successful than others, but unfortunately, none have been definitive. Attempts have been made to statistically remove the effects of halo in ratings. However, this technique does not differentiate true from illusory halo, and ideally, only illusory halo should be removed.

Traditional Model

Although there has been an enormous amount of research related to the cognitive categorization model, relatively little has been conducted to explore the traditional model as outlined by Borman (1978).

Overview. Briefly stated, this model asserts that raters can, and do, differentiate behaviors into distinct separate dimensions so that they are able to discriminately analyze observations. Borman (1978) views performance evaluation as a three-stage process: "(a) observing work-related behavior, (b) evaluating each of these behaviors, and (c) weighting these evaluations to arrive at a single rating on a performance dimension" (p. 141). Embedded in this process is an ability to remember specific ratee behaviors.

Borman (1974) alleges that individual differences in rating accuracy occur, mostly, due to raters' divergent perspectives. As an outgrowth, Borman suggests that a hybrid multi-trait-multi-rater analysis be used. This allows raters to evaluate only those dimen-

sions that they are in a good position to rate (Cascio, 1982). Across
all observers of an employee's performance, more accurate and complete
evaluations should be obtained. One implication is that raters
observing identical behaviors, perhaps those in the same organiza-
tional level, should be able to provide more consistent ratings than
those raters observing separate behaviors.

> Borman proposed that raters somehow combine the evalua-
> tions of behaviors into independent performance ratings
> on multiple dimensions. Such a model of the rating
> process implies that raters can store information in
> independent dimensional schemata and then retrieve
> this dimensionally independent information when making
> performance evaluations (Nathan & Lord, 1983, p. 103).

This, then, is a brief synopsis of two theories of cognitive
processes involved in performance appraisal. This is not to suggest
that they are the only models available, however, they do provide a
substantial foundation to further explore the mental processes under-
lying ratings of others.

Performance Patterns

What if the behavior observed is variable, as would be realis-
tically expected in organizations? Relatively few studies have
addressed this issue. Those that have, have produced inconsistent
results. The pattern of performance exhibited by an employee may
have varying effects on the ratings they subsequently receive.

In a series of experiments, Jones, Rock, Shaver, Goethals, and
Ward (1968) examined the effect that patterns of performance had on

the ascription of ability. They found that descending performers
were attributed with greater intelligence and were expected to out-
perform their ascending and random order counterparts. It appears
that performers who excel at first are viewed as more capable than
their "late bloomer" counterparts. Furthermore, this finding is
robust and replicable. Jones et al. (1968) explain that early infor-
mation about ability is heavily weighted and leads to premature
and perservering ascriptions. This seems to be consistent with the
cognitive categorization model in that early schematic impressions
are maintained regardless of subsequent patterns of performance.
However, other research findings are contradictory.

DeNisi and Stevens (1981) challenged the pattern effects by
presenting subjects with varying sales figures of a manager. One
hypothesis predicted that the manager presented as an ascending,
rather than descending, performer would receive more favorable evalua-
tions. Results supported this prediction and are in direct conflict
with the conclusion of Jones et al. (1968).

As previously mentioned, Nathan and Lord (1983) pitted the cate-
gorization model against the traditional model in a study exploring
halo in performance ratings. They found the traditional model was
generally more appropriate for explaining the cognitive processes of
performance ratings. However, they also found halo effects and evi-
dence of other errors that were unexplainable by Borman's (1978)
theory but were consistent with cognitive categorization.

The present study further examined the roles of these models in
the performance appraisal process. Innumerable researchers have

called attention to the need for further investigations in this area
(e.g., Borman, 1978, 1979; Cascio, 1982; Nathan & Lord, 1983). This
project was exploratory in nature, but should serve to enhance our
knowledge of the evaluation process.

Summary and Hypotheses

Performance appraisals can have a powerful influence on the indi-
vidual evaluated as well as on organizational effectiveness. To con-
duct evaluations of their employees, organizations most frequently
turn to the rating scale. This format, however, is plagued with in-
herent problems that create biased results.

Accuracy in the appraisal process is vital but, unfortunately,
often lacking. Research has centered on improving this state-of-
affairs by examining the affect of three primary variables: (a) the
appraisal instrument, (b) the context within which the appraisal
occurs, and (c) characteristics of the ratee and rater. Investiga-
tions concerned with the first two variables have not enlightened our
knowledge of appraisal accuracy nearly as much as the more recent
focus on ratee/rater characteristics. As part of this focus, there
have been inspections of the cognitive characteristics of raters.
One result has been the development of two theoretical models;
Feldman's cognitive categorization model and Borman's traditional
model.

Feldman's model asserts that raters rely on schemas to judge the
ratee's performance. Once a judgment is made it is likely to persist
even in the face of contradictory evidence. As a by-product of cogni-

tive categorization, halo permeates the rating process, thereby producing evaluations that do not differentiate among performance dimensions. In other words, raters perceive similarity among the components of performance and therefore tend to rate all of them favorably or unfavorably. Based on the ratee's initial behaviors, he/she will be classified into a general category, or schema, and ratings will reflect prototypical behavior associated with the early category assignment rather than based on specific observations. This will be evident not only within, but also across the rated performance dimensions.

Cognitive categorization contradicts Borman's traditional model in that Borman avers that raters evaluate behaviors independently. Reliance on global impressions is foregone. Instead, the specific performance behaviors are retrieved from memory and then averaged to arrive at a dimensional rating. Therefore, when asked to make a single rating on a component of performance, the traditional model assumes raters will average (through a weighting process) the observations within a dimension and subsequently rate performance around the mean of the performance pattern.

It is unlikely that employees will exhibit a pattern of performance that is consistent and invariable over time. For the appraisal process, this reality makes the rater's job more difficult. Assuming that the appraiser first observes behavior that is mostly favorable and later observes a primarily poor performance demonstration, what impact will this have on the subsequent ratings? If ratings are completed on unique dimensions of performance, the traditional model

predicts that raters will store observed behaviors under each dimension and average the performance. Therefore, ratings on each dimension will gravitate toward the mean of presented behaviors. Dimensional ratings will depend on the number and extremity of the behavior exhibited. On the other hand, Feldman's categorization model predicts that the ratings will reflect good performance because raters were initially exposed to positive behaviors. Consequently, they would classify the ratee as a good performer. One important implication of this model is that ratings, both within and across dimensions, will be favorable as long as the overall, initial performance is favorable. The predicted ratings would change only for Feldman's model if the pattern of observed performance was reversed. That is, if raters view primarily unfavorable behaviors initially and later see primarily favorable behaviors, Feldman's model predicts that the ratings will reflect the initial poor performance, whereas Borman's predictions remain unchanged.

To be more specific, this study investigated the following research question and hypotheses:

Research question. Which theoretical model, cognitive categorization or the traditional model, will more adequately explain the rater's cognitive processes when conducting a performance appraisal?

Assuming the superiority of the cognitive categorization model led to the following hypothesis:

Hypothesis 1a. Performance will be rated favorably if raters are presented first with predominantly good ratee behaviors and later with mostly poor performance behaviors. This will be evident both

within and across performance dimensions. Conversely, if presented first with predominantly poor behaviors and later with good behaviors, raters will evaluate performance as unfavorable.

On the other hand, if Borman's traditional model was viewed as more meritorious then a different hypothesis was necessary:

Hypothesis 1b. Across observed performances, raters will average behavior variations and the dimensional ratings will therefore reflect the mean exhibited behavior.

METHOD

This research project involved three phases: (1) the careful construction of two lecture scripts and videotapes, (2) the pretesting of the videotapes, and (3) the actual performance appraisal experiment. The first phase closely emulated the ideas and procedures used by Nathan and Lord (1983).

Construction of Videotapes

Procedure. Two videotapes were developed, both of them lasting approximately 13 minutes. For each tape, a different script was created. The same person enacted the role of a college instructor in both tapes. The lecture material was selected from the general area of communication. More specifically, the instructor discussed some of the processes of communication in organizations. Each tape, however, covered different aspects of this subject. Neither tape discussed the same or similar topics. Within each videotape the lecturer displayed four behavioral incidents in each of four performance dimensions. Based on a study by Harari and Zedeck (1973) these four dimensions comprised:

1. Organization: The lecturer's arrangement of the lecture material; the extent to which the lecturer led the class through a logical and orderly sequence of material.

2. Delivery: The lecturer's manner of conveying the lecture

material and the extent to which she used the blackboard or
audiovisual aids to clarify and emphasize important points
of her presentation; speaking and writing abilities as well
as physical mannerisms are also included.

3. Relevance: The lecturer's choice of examples used in con-
veying information; examples which were important and mean-
ingful to the audience.

4. Depth of knowledge: The lecturer's mastery of the subject
matter; this includes how well she knew the literature and
the research she reported.

One lecture script exemplified the instructor exhibiting pre-
dominantly favorable teacher behavior across all four performance
dimensions while the second script displayed primarily unfavorable
teacher behavior across the dimensions. Table 1 indicates the
specific number and type of behavioral incidents involved in each of
the tapes. A variation of good and bad incidents was used in each
videotape, rather than tapes composed of all good or all bad behaviors
because this appeared to be a more realistic representation of per-
formance observed in the real world (Borman, 1978).

Development of Lecture Scripts

As a prerequisite to videotape construction, it was imperative
to develop specific behavioral incidents that would be incorporated
into the lecture scripts.

Subjects. Forty undergraduate students at Texas A&M University
voluntarily completed a brief questionnaire.

Table 1

Number and Kind (Good/Poor) of Behavioral Incidents in Each
Performance Dimension of Each Videotape

| | Favorable Tape | | Unfavorable Tape | |
| | Incidents | | Incidents | |
Performance Dimension	Good	Poor	Good	Poor
Delivery	2	2	2	2
Organization	4	0	0	4
Relevance	2	2	2	2
Depth of Knowledge	4	0	0	4
Total:	12	4	4	12

Note. Adapted from "Cognitive Categorization and Dimensional
Schemata: A Process Approach to the Study of Halo in Performance
Ratings" by B. R. Nathan and R. G. Lord, 1983, Journal of Applied
Psychology, 68, p. 106. Copyright 1983 by the American Psycho-
logical Association. Adapted by permission.

Procedure. Fifty specific examples characterizing both favorable and unfavorable behaviors from the four performance dimensions were developed by the authors. The following statement exemplified a favorable behavior in the organizational dimension: "At the end of the lecture, the instructor summarized what was discussed". A bad behavioral incident in the depth of knowledge dimension was depicted by: "The lecturer failed to remember the results of a research study".

Subjects, provided with a complete set of definitions and descriptions of the performance dimensions, were asked to review and critique the behavioral incidents. They were requested to (a) assign each incident to the dimension they believed it represented, and (b) indicate whether they felt it was an example of poor, average, or excellent behavior through the use of a seven-point Likert scale. The two lecture scripts were constructed from the behavioral incidents found to be most indicative of good or bad performance within each dimension. To be considered acceptable for inclusion in the scripts, an item reflecting good (poor) instructor behavior had to be rated a six or seven (one or two) by more than 50% of the respondents. Additionally, more than 50% of the respondents needed to agree on their assignment of the item to a particular dimension.

Videotape Pretesting

Subjects. Thirty undergraduate students (14 males, 16 females) participated in the pretesting to fulfill an introductory psychology course requirement. Students involved in the development of the

lecture scripts could not participate in this phase of the experiment.

Procedure. The two videotaped lectures were shown to all participants. After each tape, three 7-point Likert rating scales (with extreme anchors of "very poor" and "very good") were provided and subjects were requested to rate the instructor's performance in each separate tape. Results indicated that the tape designed to exhibit predominantly favorable behaviors received a mean rating of 4.7 (corresponding to the scale anchor of "somewhat good") while the tape with primarily unfavorable behaviors received a mean rating of 1.8 (corresponding to the scale anchor of "mostly poor"). This appeared to be an adequate differentiation of the two tapes to permit their use in the next phase of the experiment, $t(58) = 12.29$, $p < .001$.

Performance Appraisal Experiment

Subjects. Fifty-four undergraduate students (30 males, 24 females) participated in exchange for course credit. They were randomly assigned to one of two experimental conditions. The average age of participants was 19.5 years; 63% were freshmen; and 39% were in a business related field of study. Subjects involved in either the earlier development of the lecture scripts or the pretesting of the videotapes could not participate in this experiment.

Procedure. At the onset of the study subjects were given a complete explanation of each performance dimension and were also informed that they would later appraise the performance of a college instructor based on these dimensions. Nathan and Lord (1983) provided the rationale for following these procedures:

> First, they embody many of the steps that are commonly
> thought to produce good ratings. Second, they should
> produce controlled rather than automatic processing
> of information (Feldman, 1981). Third, if the tra-
> ditional model is correct, such conditions should
> minimize the occurrence of halo, leading to differen-
> tial predictions for the traditional as compared to
> the categorization model (p. 105).

Subjects were then shown the two videotaped lectures, however, the tapes were presented one day apart and during the final, third day session they were given the seven-point scales on which their ratings of the instructor were made.

Independent variable. As previously mentioned, the single independent variable was the order of videotape presentation. One group of subjects viewed the tapes in descending performance order. That is, they viewed the favorable tape during the first session and the unfavorable tape during the second session. The order of presentation was reversed for the remaining group. By showing the tapes with a one day temporal delay and manipulating the order of presentation, it was believed that a more realistic replication of the kind of observations made in the business world would be reflected. Typically, performance appraisers do not observe individuals always behaving consistently throughout the entire rating period.

Dependent variable. After viewing both tapes, participants assessed the instructor's overall performance as well as her specific performance in each of the four dimensions. These judgments were measured on seven-point Likert rating scales. Three ratings questioning the lecturer's preparation and presentation of the lecture material were combined to form the organization scale.

Three questions pertaining to the lecturer's speaking and writing abilities as well as the use of visual aids were combined to form a measure of the lecturer's delivery. A relevance scale was developed by combining three questions covering the lecturer's choice of meaningful and interesting examples. Finally, a measure of the lecturer's depth of knowledge was formed by combining three questions relating to the lecturer's mastery of the subject matter and understanding of the literature and research reported. The reliabilities of the organization, delivery, relevance, and depth of knowledge scales were computed, revealing a coefficient alpha equal to .79, .53, .75, and .78 respectively.

Additionally, subjects were presented with a recognition memory test consisting of a list of 32 critical incidents. Of these incidents, eight were selected from the favorable tape, eight from the unfavorable tape, and the remaining 16 incidents were not present in either videotape. Subjects indicated whether or not each incident appeared in either tape. In this way any differences in the memory capabilities of each experimental group could be detected. Furthermore, the model that can best explain the rater's cognitive processes should be supported, in part, by the responses on the recognition measure. Borman's model would suggest that raters accurately recall the behaviors exhibited by the instructor. However, Feldman's model would predict less accurate recognitions. More specifically, the rater's recall would be based on the schema activated after viewing the first videotape rather than on each specific behavior exhibited.

Manipulation check. To ensure the efficacy of the independent variable, subjects responded to a brief post-experimental questionnaire. This measured the extent to which participants viewed each tape as a favorable or unfavorable demonstration of lecture performance.

RESULTS

The results of this study are presented in three sections. These include the manipulation check, the performance ratings, and a memory recognition analysis.

Manipulation Check

In order to assess the extent to which the experimental manipulation was effective, respondents completed three questions: (1) specifically in the first videotape viewed, how would you evaluate the lecturer's performance, (2) specifically in the second videotape viewed, how would you evaluate the lecturer's performance, and (3) considering both videotapes viewed, in which one was the lecturer's performance most favorable. Responses to the first two questions were collected on seven-point rating scales. The third question required subjects to check either the first videotape viewed or the second (dummy coded a 1 and 2 respectively). Summary descriptive statistics along with the results of a one-way analysis of variance for each of the three manipulation check items are presented in Table 2. In reviewing the results, it is necessary to remember that participants in group 1 viewed the favorable lecture first, followed by the presentation of the unfavorable lecture during the second experimental session. The order of tape presentation was reversed for group 2.

The analysis of variance for the first item indicates a significant difference in the groups' perceptions of the first videotape,

Table 2

Analysis of Variance and Descriptive Statistics for the Manipulation
Checks

Videotape	Group 1	Group 2	Total	F^a	eta^2
Tape 1					
M	3.89	1.22	2.56	126.92*	.71
SD	1.15	.42	1.60		
Tape 2					
M	1.89	4.22	3.06	46.99*	.47
SD	.80	1.58	1.71		
Both Tapes					
M	1.04	1.96	1.50	312.50*	.86
SD	.19	.19	.50		

Note. Tape order for Group 1 was favorable/unfavorable and tape
order for Group 2 was unfavorable/favorable.

$^a\underline{df} = 1,52$

*$\underline{p} < .001$

$F(1,52) = 126.92$, $p < .0001$. This difference was also found in the perceptions of the second videotape viewed by each group, $F(1,52) = 46.99$, $p < .0001$.

In order to correctly and fully test the competing hypotheses it was necessary for participants to perceive both a favorable and an unfavorable lecture performance. The manipulation check indicates that this requisite was not met. Most significant is the finding that the lecturer's performance in the "favorable" tape was perceived as average. Evidently, the performance ratings provided by participants in the pretest experiment did not coincide with the perceptions of the respondents in this phase. Unfortunately, this result must color the interpretation of the data reported subsequently. Any hypotheses testing that relies on the presence of a favorable and unfavorable tape (rather than an average and an unfavorable tape) may be uninterpretable or, at least, inappropriate. However, tests of hypotheses that are merely dependent on a differentiation of the tapes' favorabilities may be salvageable.

The analysis of the third manipulation check item revealed that participants easily identified which tape reflected the more favorable lecture performance, $F(1,52) = 312.50$, $p < .0001$. A mean of 1.0 was expected for group 1, indicating that of the two tapes viewed, the first was perceived as more favorable. On the other hand, a mean of 2.0 was expected for group 2, indicating that the second tape viewed was selected as the better performance. The actual means for groups 1 and 2 were 1.04 and 1.96 respectively.

Also included in the post-experimental questionnaire were three

items designed to assess the participants' (1) prior experience in evaluating an instructor's performance, (2) knowledge of the lecture topic (communication in organizations), and (3) familiarity with the female lecturer in the tapes. Responses (all coded yes = 1, no = 2) indicated that there were no significant differences in (1) the groups' prior experience with evaluating an instructor's performance, $F(1,52) = 1.17$, $p > .28$, with means of 1.56 and 1.41 for groups 1 and 2 respectively, or (2) their knowledge of the lecture topic, $F(1,52) = 0.11$, $p > .74$ with means of 1.82 and 1.78 respectively. Additionally, no participants in either group indicated that they had prior familiarity with the lecturer in the videotapes.

Performance Ratings

Rating scales. Before examining the actual performance ratings a preliminary check was conducted to determine the degree to which the four dimensional scales were intercorrelated. Table 3 presents these results. By examining the correlations, the degree of halo within the ratings can be assessed. Higher correlations indicate less rater discrimination of the lecturer's behaviors and thus more halo (Saal, Downey, & Lahey, 1980). Overall, halo appears to be prominent in the ratings. The correlation between the organization and delivery dimensions ($r = .75$, $p < .001$) is particularly high. As Feldman's model predicts, there was a reliance on global impressions. The presence of halo contradicts Borman's prediction that subjects would differentiate performance behaviors rather than deperd on vague abstractions.

Table 3

Intercorrelations of Performance Dimension Scales

Dimensional Scale	1	2	3	4
1. Organization	1.00	--	--	--
2. Delivery	.75*	1.00	--	--
3. Relevance	.60*	.69*	1.00	--
4. Depth of Knowledge	.56*	.62*	.59*	1.00

Note. N = 54

*$p < .001$

As previously stated, if Borman's model is correct the perfor-
mance ratings should reflect the mean of the instructor's behaviors.
There should not be any significant differences in the ratings pro-
duced by the two experimental groups. However, if Feldman's model
is supported, the ratings should mirror the videotape initially viewed
by the ratee. Accordingly, significant differences would be expected
in the ratings of the two groups. The means and standard deviations
for the four specific scale ratings are presented in Table 4. This
analysis shows that each group provided similar mean dimensional
ratings.

Furthermore, the effects of the experimental manipulation were
assessed through a multivariate analysis of variance with the four
performance scales combined to serve as the dependent variable,
while the order of tape presentation served as the independent
variable. Wilks' lambda revealed an $\underline{F}$(approx) = .40, $\underline{df}$ = 4,49, $\underline{p}$
> .80. Clearly, across all dependent variables there were no sig-
nificant differences in the ratings of the two experimental groups.
With these results, statistical guidelines did not permit a subse-
quent series of univariate, one-way analyses of variance with each
performance scale serving as a separate dependent variable.

As previously stated, the reliability of the delivery dimensional
scale was .53, which was particularly low. Therefore, each item
comprising the delivery scale was analyzed separately. The analysis
of variance, however, did not reveal any significant differences.

Finally, on a seven-point rating scale, subjects were asked to
evaluate the lecturer's overall performance across both videotapes.

Table 4

Descriptive Statistics for Performance Dimension Scales

Dimensional Scale	Group 1	Group 2	Total Sample
Organization			
M	3.14	3.27	3.20
SD	.93	.97	.94
Delivery			
M	3.07	3.02	3.05
SD	.72	.91	.81
Relevance			
M	3.15	3.17	3.16
SD	.90	1.09	.99
Depth of Knowledge			
M	3.19	3.32	3.25
SD	.90	.88	.88

These responses served as the dependent variable in a one-way analysis of variance while the two experimental groups served as the independent variable. The analysis revealed no significant differences in the overall performance ratings, $F(1,52) = .02$, $p > .88$. The mean overall rating for group 1 was 2.56 while for group 2 the mean was 2.59. These ratings are nearly an exact average of the groups' perceptions of the two tapes separately (see Table 2, p. 29). This result partially supports Borman's predictions, in that subjects appeared to base their overall rating on an average of the individual tape ratings rather than rely on the schema activated by the initial tape viewed, as predicted by Feldman. However, due to the perceptions of the "favorable" tape as only average, Feldman's predictions could not be appropriately tested.

Memory Recognition

Memory check. To determine whether there were any differences in the memory capacities of the two experimental groups, a correlated t-test was computed. This analysis examined the subjects' ability to accurately distinguish the behavioral items that were present from those that were absent in the videotapes. Due to the random assignment of subjects to experimental conditions, differences in memory were neither expected nor found, $t(53) = .36$, $p > .72$.

Design. The 32 behavioral items were grouped according to their (a) association with the favorable or unfavorable tape, (b) presence or absence in either videotape, and (c) exemplification of good or bad lecturer behavior. The design for this part of the

analyses was therefore a 2 (order of tape presentation) X 2 (favorable/
unfavorable videotape) X 2 (presence/absence of item) X 2 (good/bad
behavior). Due to perceptions of the favorable tape as merely average,
this design was subsequently altered to a 2 (tape order) X 2 (presence/
absence) X 2 (good/bad) analysis with one between subjects factor
and two within subjects factors.

Item analyses. The means and cell deviations associated with
the significant findings of an analysis of variance are presented in
Table 5. Additionally, all results of this repeated measures analysis
of variance are presented in Table 6. Each recognition item was
coded zero if the response was inaccurate. Otherwise, if the subject
accurately identified the item, it was given a code of one. The
grand mean across all memory items was .66 which is not particularly
high. This finding does not support Borman's contention that raters
remember the specific behaviors exhibited by the ratee.

Table 6 reveals a marginally significant main effect for the
order of tape presentation. This finding, coupled with the signifi-
cant interaction of tape order by presence/absence of item, indicates
a probable difference in the memory abilities of the two groups.
More specifically, the means in Table 5 show that both groups
were equivalently accurate in recalling the present items, however,
there was a large difference (.60 for group 1 and .72 for group 2)
in their abilities to accurately recognize items that were not present
in either tape.

The most significant result was the interaction between the
presence/absence of items and their exemplification of a good/bad

Table 5

<u>Means and Cell Deviations for Recognition</u>
<u>Memory Items (Significant Results Only)</u>

	Tape Order	
	Group 1	Group 2
M	.64	.69
SD	.20	.17

Presence/Absence of Items	Tape Order	
	Group 1	Group 2
Present Items		
M	.68	.66
SD	.19	.15
Absent Items		
M	.60	.72
SD	.21	.20

Presence/Absence of Items	Behavior	
	Good	Bad
Present Items		
M	.63	.71
SD	.18	.16
Absent Items		
M	.72	.60
SD	.17	.25

<u>Note</u>. Higher mean values indicate more accurate recognition.

Table 6

Analyses of Variance for Recognition Memory Items

Source	MS	F	eta^2
Tape Order (T)	.15	3.84+	.05
Presence/Absence of Item (P)	.01	.14	.00
Good/Bad Behavior (G)	.02	.67	.01
T X P	.29	6.95*	.09
T X G	.00	.03	.00
P X G	.57	14.46**	.18
T X P X G	.09	2.24	.03

Note. df = 1,52

+p < .10

*p < .05

**p < .001

behavior, $F(1,52) = 14.46$, $p < .0005$. The means indicate that subjects more accurately recalled items that were present in the tapes and examples of bad behaviors than they recognized items that were absent and good. This interaction lends support to Feldman's prediction that subjects rely on schemas rather than on specific behaviors to evaluate performance. The schema in this case would reflect an instructor exhibiting predominantly poor performance across both videotapes, as indicated in the overall performance ratings. Therefore, cognitive categorization predicts that responses should reflect a global recognition of items exemplifying poor behaviors and a global lack of recognition of items exemplifying favorable performance. This pattern is reflected in the lower recognition accuracy for good/present items and bad/absent items and at the same time a higher accuracy for recognizing good/absent items and bad/present items. Subjects more accurately remembered behaviors that reinforced their schema of an incompetent instructor and failed to accurately recall the behaviors that were inconsistent with the schema.

DISCUSSION

Performance ratings can have a pervasive impact on many facets of the organization. Gaining a better understanding of the rater's implicit cognitive processes was the primary aim of this study. Specifically, it was designed to assess which of two theoretical models most adequately explains these cognitive operations. The original research question focused on two competing hypotheses. Assuming the superiority of one cognitive model over the other, each hypothesis made corresponding predictions of the performance appraisal outcomes. Based on the analytical results, it seems that neither the traditional nor categorization model alone can completely account for the cognitions of the subjects involved in the present experiment. Both models are partially supported and at the same contradicted.

A brief review of the results indicates support for Borman's traditional model in two fundamental areas, both of which directly relate to the performance ratings. First, there were no significant differences in the groups' ratings on the four dimensional scales. Secondly, evaluations of the lecturer's overall performance did not significantly differ across the groups. Instead these ratings centered around the mean of all the lecturer's behaviors. According to Borman, there should not be differences in the scale and overall ratings because raters accurately remember specific dimensional behaviors. Based on the findings reported in the following section, this assumption appears to be erroneous.

The cognitive categorization model also finds partial support in the data. Unlike the traditional model, however, this support does not generally stem from the performance ratings but rather is embedded in the results of the recognition memory task. In contrast to Borman's predictions, the analysis revealed an overall inability of raters to accurately recall the behavioral items. This is inconsistent with the accuracy observed in the performance ratings. Subjects were able to accurately rate the instructor's performance across both videotapes, as witnessed in the mean overall ratings. However, the recognition task revealed that subjects did not remember the specific behaviors that collectively produced the overall performance. Feldman's model predicts that recall should be based on the schema initially activated rather than on specific lecturer behaviors. Contrary to experimental intentions, the schema activated by viewing both videotapes was predominantly of an unfavorable lecturer. Therefore, the categorization model contends that raters' recognition scores should indicate a global recall of items reflecting unfavorable lecture behaviors. In completing the recognition task, subjects should agree to seeing more unfavorable than favorable behaviors due to reliance on their schemas. This should generally occur across all recognition items irrespective of their actual presence or absence in the videotapes. In effect, saying "yes" to the unfavorable items leads to higher accuracy scores for those items that were actually present but lower accuracy for absent items. Conversely, reliance on the unfavorable lecturer schema leads to negative responses to favorable items, thereby producing higher accuracy for absent favorable

items and lower accuracy for present items. This pattern is clearly revealed in the memory recognition scores.

The halo present in the ratings further supports categorization's prediction that raters classify people into global categories. This general and unfavorable impression of the lecturer thereby produced a large halo effect.

Unfortunately, a complete test of the competing hypotheses could not be conducted due to perceptions of the favorable tape as merely average. In an attempt to mimic real world performance, each videotape contained both favorable and unfavorable behaviors. This may have precluded raters from perceiving the favorable tape more positively. Perhaps for exploratory purposes there needs to be a greater disparity in the good and bad behavioral examples. Furthermore, there may have been an inherent problem in the use of videotapes. This is not characteristic of raters in the organization where there is opportunity for dynamic interaction with the ratee. The sterile tapes provided no such interchange. One potential solution to this problem is the addition of a fifth performance dimension in the videotapes, namely, interpersonal relations with students (Harari & Zedeck, 1973). A more favorable and holistic impression may develop by viewing an interaction with others.

How does one account for the general finding that Borman's model is most applicable when explaining the performance ratings while Feldman's model best explains the results of the recognition memory task? Perhaps raters utilize different cognitive strategies at different levels of appraisal. Fiske and Taylor (1984) outline four

levels of specificity that may create variations in memory accuracy. Recall can be more or less accurate about (1) people in general, (2) specific people, (3) specific attributes, or (4) specific attributes for specific people. Participants in the present study were asked to make general performance ratings that merely required reliance on a global impression. They were also requested to recall specific behavioral items, thus requiring a more accurate encoding process and efficient search of memory. In terms of the four levels of Fiske and Taylor (1984), the performance ratings depend on raters' accuracy about specific people, in this case one specific lecturer. On the other hand, accuracy on the recognition items relies on memory about specific attributes for specific people.

Fiske and Tyalor (1984) also argue that efficiency and accuracy are often forfeited for each other. This may explain the low levels of accuracy found in the recognition items in this study. It is not efficient for raters to categorize information at a specific behavioral level when they can be relatively accurate on performance ratings by simply depending on a more general, summary evaluation of the target stimulus. Thus, in order for subjects to efficiently store and access information for the performance ratings, some accuracy was sacrificed, as evidenced in the recognition task.

In a related vein, Lord (1985) suggests that there are three approaches to defining accuracy in behavioral measurement. These include behavioral accuracy, classification accuracy, and differences in rater decision criteria. The first two approaches are of particular interest to the findings of the current research project. According

to Lord (1985), high behavioral accuracy occurs when raters process behavioral information effectively and, at the same time, distinguish this information from behaviors that were not observed but appear plausible based on knowledge about the ratee. This definition corresponds to the type of accuracy required to correctly respond to the behavioral recognition items in the present study.

Classification accuracy refers to the categorization of information based on a global impression of the ratee. The target stimulus is simplified and there is merely an overall classification of information. When completing the performance ratings, participants in this study only had to rely on classification accuracy. This approach enabled subjects to combine efficiency and accuracy to their greatest advantage.

The findings of this project are generally consistent with the results of the experiment conducted by Nathan and Lord (1983). They found that predictions of the traditional model were upheld in their performance ratings. However, the existence of halo in the ratings, and errors in recognizing specific behavioral incidents lent support to categorization. Together, these studies suggest that future research is needed to uncover the generalizability of the results to the organization. Both experiments presented videotaped stimuli to subjects in a controlled environment. Performance appraisals, however, are often made under more complex, confounded, and ambiguous circumstances.

Additionally, the ability of these findings to generalize is inhibited by the use of upward rather than downward appraisals.

Research needs to focus on supervisors' ratings of subordinates as that is the normal organizational procedure.

R.   ·rch is also needed that would span a longer time period. Although this project examined cognitive processes across five days, this is a minimal demand on memory in comparison to the yearly appraisals that are common in organizations. In this way, the effect of performance patterns can be demonstrated. Murphy, Balzer, Lockhart, and Eisenman (1985) recently explored the impact of prior performance on evaluations of present performance. In part they found that deviations from previous performance patterns can intensify the difficulty of making accurate appraisals. A strong contrast effect was observed in both the performance evaluations and the ratings of the frequency of numerous critical behaviors. Unfortunately, they were unable to adequately explain the memory processes involved, in spite of conditions that minimized memory requirements. A synthesis of this focus on performance patterns together with a longitudinal cognitive emphasis should reveal unique applications to appraisals in organizational settings.

Finally, performance appraisal models encompassing the complex cognitive processes of raters are needed to generate and advance research. In this respect, the model and research propositions developed by DeNisi, Cafferty, and Meglino (1984) provide a promising foundation on which future appraisal research can be based.

REFERENCES

Bernardin, H. J., & Cardy, R. L. (1982). Appraisal accuracy: The ability and motivation to remember the past. Public Personnel Management, 11(4), 352-357.

Bernardin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Bigoness, W. J. (1976). Effect of applicant's sex, race, and performance on employers' performance ratings: Some additional findings. Journal of Applied Psychology, 61(1), 80-84.

Borman, W. C. (1974). The rating of individuals in organizations: An alternate approach. Organizational Behavior and Human Performance, 12, 105-124.

Borman, W. C. (1978). Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 63(2), 135-144.

Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64(4), 410-421.

Cantor, N. & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 35(1), 38-48.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 12). New York: Academic Press.

Cascio, W. F. (1982). Applied psychology in personnel management (2nd ed.). Reston, VA: Reston.

Cooper, W. H. (1981). Ubiquitous halo. Psychological Bulletin, 90(2), 218-244.

Cooper, W. H. (1983). Internal homogeneity, descriptiveness, and halo: Resurrecting some answers and questions about the structure of job performance rating categories. Personnel Psychology, 36(3), 489-502.

DeCotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. Academy of Management Review, 3(3), 635-646.

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: A model and research propositions. Organizational Behavior and Human Performance, 33(3), 360-396.

DeNisi, A. S., & Stevens, G. E. (1981). Profiles of performance, performance evaluations, and personnel decisions. Academy of Management Journal, 24(3), 592-602.

Dunnette, M. D., & Borman, W. C. (1979). Personnel selection and classification systems. Annual Review of Psychology, 30, 477-525.

Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66(2), 127-148.

Fiske, S. T., & Taylor, S. E. (1984). Social Cognition. Reading, MA: Addison-Wesley.

Foti, R. J., Fraser, S. L., & Lord, R. G. (1982). Effects of leader-
ship labels and prototypes on perceptions of political leaders.
Journal of Applied Psychology, 67(3), 326-333.

Fox, S., Bizman, A., & Herrmann, E. (1983). The halo effect: Is it
a unitary concept? Journal of Occupational Psychology, 56(4),
289-296.

Harari, O., & Zedeck, S. (1973). Development of behaviorally anchored
scales for the evaluation of faculty rating. Journal of Applied
Psychology, 58, 261-265.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process
focus. In L. L. Cummings & B. M. Staw (Eds.), Research in
organizational behavior (Vol. 5). Greenwich, CT: JAI Press.

Johnson, D. M. (1963). Reanalysis of experimental halo effects.
Journal of Applied Psychology, 47(1), 46-47.

Johnson, D. M., & Vidulich, R. N. (1956). Experimental manipulation
of the halo effect. Journal of Applied Psychology, 40(2), 130-134.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M.
(1968). Pattern of performance and ability attribution: An
unexpected primacy effect. Journal of Personality and Social
Psychology, 10(4), 317-340.

Kane, J. S., & Lawler, E. E. III. (1979). Performance appraisal
effectiveness: Its assessment and determinants. In B. M. Staw
(Ed.), Research in organizational behavior (Vol. 1). Greenwich,
CT: JAI Press.

Kenny, D. A., & Berman, J. S. (1980). Statistical approaches to the

correction of correlational bias. Psychological Bulletin, 88, 288-295.

Kulik, J. A. (1983). Confirmatory attribution and the perpetuation of social beliefs. Journal of Personality and Social Psychology, 44(6), 1171-1181.

Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87(1), 72-107.

Landy, F. J., & Farr, J. L. (1983). The measurement of work performance: Methods, theory, and applications. New York: Academic Press.

Latham, G. P., & Wexley, K. N. (1980). Improving performance through effective performance appraisal. Reading, MA: Addison-Wesley.

Lingle, J. H. (1979). Probe recognition speed as a measure of thought activation during memory-based impression judgments. Dissertation Abstracts International, 39(8-B), 4107.

Lingle, J. H., & Ostrom, T. M. (1979). Retrieval selectivity in memory-based impression judgments. Journal of Personality and Social Psychology, 37(2), 180-194.

Lord, R. G. (1985). Accuracy in behavioral measurement: An alternative definition based on raters' cognitive schema and signal detection theory. Journal of Applied Psychology, 70(1), 66-71.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 37(11), 2098-2109.

Murphy, K. R., Balzer, W. K., Lockhart, M. C., & Eisenman, E. J. (1985). Effects of previous performance on evaluations of present performance. Journal of Applied Psychology, 70(1), 72-84.

Nathan, B. R., & Lord, R. G. (1983). Cognitive categorization and dimensional schemata: A process approach to the study of halo in performance ratings. Journal of Applied Psychology, 68(1), 102-114.

Ostrom, T. M., Lingle, J. H., Pryor, J. B., & Geva, N. (1980). Cognitive organization of person impressions. In R. Hastie, T. M. Ostrom, E. B. Ebbesen, R. S. Wyer, Jr., D. L. Hamilton, & D. E. Carlston (Eds.), Person memory: The cognitive basis of social perception. Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.

Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8, 382-439.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin, 88(2), 413-428.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences pf person perception. Journal of Personality and Social Psychology, 36(9), 941-950.

Srull, T. K., & Wyer, R. S., Jr. (1979). The role of category accessibility in the interpretation of information about persons:

Some determinants and implications. _Journal of Personality and Social Psychology_, _37_(10), 1660-1672.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), _Social cognition: Vol. 1. The Ontario symposium_. Hillsdale, NJ: Lawrence Erlbaum.

Thorndike, E. L. (1920). A constant error in psychological ratings. _Journal of Applied Psychology_, _4_, 25-29.

Vance, R. J., Winne, P. S., & Wright, E. S. (1983). A longitudinal examination of rater and ratee effects in performance ratings. _Personnel Psychology_, _36_(3), 609-620.

Warmke, D. L., & Billings, R. S. (1979). Comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. _Journal of Applied Psychology_, _64_, 124-131.

Zedeck, S., & Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. _Journal of Applied Psychology_, _67_(6), 752-758.

442GP DISTRIBUTION LIST

LIST 1
MANDATORY

Defense Technical Information Center   (12 copies)
ATTN: DTIC DDA-2
Selection and Preliminary Cataloging Section
Cameron Station
Alexandria, VA   22314

Library of Congress
Science and Technology Division
Washington, D.C.   20540

Office of Naval Research   (3 copies)
Code 442GP
800 N. Quincy Street
Arlington, VA   22217

Naval Research Laboratory   (6 copies)
Code 2627
Washington, D.C.   20375

Office of Naval Research
Director, Technology Programs
Code 200
800 N. Quincy Street
Arlington, VA   22217

Psychologist
Office of Naval Research
Detachment, Pasadena
1030 East Green Street
Pasadena, CA   91106

Deputy Chief of Naval Operations
    (Manpower, Personnel, and Training)
Head, Research, Development, and
    Studies Branch (OP-01B7)
1812 Arlington Annex
Washington, DC   20350

Director
Civilian Personnel Division (OP-14)
Department of the Navy
1803 Arlington Annex
Washington, DC   20350

Deputy Chief of Naval Operations
    (Manpower, Personnel, and Training)
Director, Human Resource Management Division
(OP-15)
Department of the Navy
Washington, DC   20350

Chief of Naval Operations
Head, Manpower, Personnel, Training
    and Reserves Team (Op-964D)
The Pentagon, 4A478
Washington, DC   20350

Chief of Naval Operations
Assistant, Personnel Logistics
    Planning (Op-987H)
The Pentagon, 5D772
Washington, DC   20350

NAVMAT

Program Administrator for Manpower,   MANPOWER REPORTS ONLY
Personnel, and Training
MAT-0722
800 N. Quincy Street
Arlington, VA 22217

Naval Material Command
Management Training Center
NAVMAT 09M32
Jefferson Plaza, Bldg #2, Rm 150
1421 Jefferson Davis Highway
Arlington, VA   20360

Naval Material Command
Deputy Chief of Naval Material, MAT-03
Crystal Plaza #5
Room 236
Washington, DC   20360

Naval Personnel R&D Center   (4 copies)
Technical Director
Director, Manpower & Personnel
  Laboratory, Code 06
Director, System Laboratory, Code 07
Director, Future Technology, Code 04
San Diego, CA   92152-6800

Navy Personnel R&D Center
Washington Support Office
Ballston Tower #3, Room 171
Arlington, VA   22203-1923
E

Naval Hospital
Psychology Department
San Diego, CA 92134

Commanding Officer
Naval Submarine Medical
    Research Laboratory
Naval Submari Base
New London, Bo 900
Groton, CT 06 9

Commanding Officer
Naval Aerospace Medical
    Research Lab
Naval Air Station
Pensacola, FL 32508

Naval Medical R&D Command
Program Manager for Human
  Performance (Code 404)
National Naval Medical Center
Bethesda, MD 20014

Wilkins Biomedical Library
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92138-9174

Naval Postgraduate School   (3 copies)
ATTN: Chairman, Dept. of
  Administrative Science
Department of Administrative Sciences
Monterey, CA 93940

U.S. Naval Academy
ATTN: Chairman, Department
  of Leadership and Law
Stop 7-B
Annapolis, MD   21402

Superintendent
ATTN: Director of Research
Naval Academy, U.S.
Annapolis, MD   21402

Commanding Officer
Organizational Effectiveness Center
Naval Training Center
San Diego, CA 92133-9000

Commanding Officer
Organizational Effectiveness Center
Naval Submarine Base New London
P.O. Box 81
Groton, CT 06349

Commanding Officer
Organizational Effectiveness Center
Naval Air Station
Mayport, FL 32228

Commanding Officer
Organizational Effectiveness Center
Pearl Harbor, HI 96860

Commanding Officer
Organizational Effectiveness Center
Naval Base (Bldg. NH-46)
Charleston, SC 29408

Commanding Officer
Leadership & Organizational Effectiveness
  School
Naval Air Station Memphis
Millington, TN 38054-5099

Commanding Officer
Organizational Effectiveness Center
1300 Wilson Boulevard, rm 114A8
Arlington, VA 22209

Commanding Officer
Organizational Effectiveness Center
5621-23 Tidewater Drive
Norfolk, VA 23509

Commander
Organizational Effectiveness Center
5621 Tidewater Drive
Norfolk, VA 23509

Commanding Officer
Organizational Effectiveness Center
Naval Air Station Whidbey Island
Oak Harbor, WA 98278-9000

Commanding Officer
Organizational Effectiveness Center
Box 23
FPO New York 09510

Commanding Officer
Organizational Effectiveness Center
Box 60
FPO San Francisco 96651

Commanding Officer
Organizational Effectiveness System, Pacific
Pearl Harbor, HI 96860

Commanding Officer
Organizational Effectiveness System, Atlantic
5621 Tidewater Drive
Norfolk, VA 23509

Commanding Officer
U.S. Navy Organizational Effectiveness System, Europe
FPO New York 09510

Commanding Officer
U.S. Navy Organizational Effectiveness Center
Box 4
FPO Seattle 98762-2920

Naval Military Personnel Command   (2 copies)
HRM Department (NMPC-6)
Washington, DC   20350

Commander
Naval Training Equipment Center
(Code 1 - Resource Center)
Orlando, FL   32813

Commanding Officer
ATTN: TIC, Bldg. 2068
Naval Training Equipment Center
Orlando, FL   32813

Chief of Naval Education & Training (N-22)
Naval Air Station
Pensacola, FL   32508

Chief of Naval Technical Training
ATTN: Code D17
NAS Memphis (75)
Millington, TN   38D54

Navy Recruiting Command
Director, Recruiting Advertising Dept.
Code 43
801 North Randolph Street
Arlington, VA   22203

Naval Weapons Center
Code 094
China Lake, CA   93555

Dr. Eduardo Salas
Human Factors Division (Code 712)
Navy Training Systems Center
Orlando, FL 32813-7100

Headquarters, U.S. Marine Corps
Code MPI-20
Washington, DC 20380

Headquarters, U.S. Marine Corps
ATTN: Scientific Adviser,
    Code RD-1
Washington, DC 20380

Director
Education Center (E 032B)
MCDEC
Quantico, VA 22134-5050

Commanding Officer
Education Center (E031)
MCDEC
Quantico, VA 22134

Marine Corps Command and
  Staff College
Education Center
Quantico, VA 22134

Defense Advanced Research
  Projects Agency
Director, Cybernetics
  Technology Office
1400 Wilson Blvd, Rm 625
Arlington, VA 22209

Professor Douglas E. Hunter
Defense Intelligence School
Washington, DC 20374-6111

Dr. Brian Usilaner
GAO
Washington, DC 20548

School Management Unit
National Institute of Education
1200 19th Street, N.W.
Mail Stop 17
Washington, DC 20208

National Institute of Mental Health
Division of Extramural Research Programs
5600 Fishers Lane
Rockville, MD 20852

Information Analyst
Center for Studies of Minority Group
  Mental Health
Parklawn Building, Rm 11-94
5600 Fishers Lane
Rockville, MD 20857

Chief, Personnel Policy Analysis Branch
U.S. Coast Guard (G-P-1/2)
Washington, D.C. 20593

Social and Developmental Psychology
  Program
National Science Foundation
Washington, D.C. 20550

Dr. Earl Potter
Department of Economics & Management
U.S. Coast Guard Academy
New London, CT 06320

Douglas B. Blackburn, Director
National Defense University
Mobilization Concepts Development
  Center
Washington, D.C. 20319

Chairman, Dept. of Medical Psychology
School of Medicine
Uniformed Services University of
  the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814

LIST 10
ARMY

Headquarters, FORSCOM
ATTN: AFPR-HR Ltc. Sellards
Ft. McPherson, GA 30330

Army Research Institute
Field Unit - Ft. Leavenworth
P.O. Box 290
Leavenworth, TX 66048

Technical Director   (3 copies)
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Head, Department of Behavior
Science and Leadership
U.S. Military Academy, New York 10996

LTC. Frederick J. Manning
Deputy Director
Division of Neuropsychiatry
Walter Reed Army Institute
Washington, DC 20307-5100

Army Military Personnel Command
Attn: DAPC-OE
200 Stovall Street
Alexandria, VA 22322

Army Research Institute
Attn: PERI-SF (Mr. Dennis Leedom)
5001 Eisenhower Avenue
Alexandria, VA 22333

Commandant
USA OECS
Attn: ATXW-RMA-S
Ford Ord, CA 93941-7300

Air University Library
LSE 76-443
Maxwell AFB, AL 36112

Head, Department of Behavioral
Science and Leadership
U.S. Air Force Academy, CO 80840

Major Robert Gregory
USAFA/DFBL
U.S.A.F. Academy
Colorado Springs, CO 80840-5941

A. R. Fregley
AFOSR/NL
Building 410
Bolling Air Force Base
Washington, DC 20332-6448

Technical Director
AFHRL/MO(T)
Brooks AFB
San Antonio, TX 78235

AFMPC/MPCYPR
Randolph AFB, TX 78150

LIST 12
MISCELLANEOUS

Australian Embassy
Office of the Air Attache (S3B)
1601 Massachusetts Avenue, N.W.
Washington, D.C. 20036

British Embassy
Scientific Information Office
Room 615
3100 Massachusetts Avenue NW
Washington, DC   20008

Canadian Defense Liaison Staff,
    Washington
ATTN: CDRD
2450 Massachusetts Avenue, N.W.
Washington, DC   20008

Commandant, Royal Military
College of Canada
ATTN: Department of Military
Leadership and Management
Kingston, Ontario   K7L 2W3

National Defense Headquarters
ATTN: DPSRSC
Ottawa, Ontario K1A OK2

LIST 13

CURRENT CONTRACTORS

Dr. Robert Archer
Eastern Virginia Medical Authority
P.O. Box 1980
Norfolk, VA 23501

Dr. Janet Barnes-Farrell
University of Connecticut
Department of Psychology
University of Connecticut Research Foundation
Storrs, CT 06268

Dr. Lee Roy Beach
University of Washington
1107 N.W. 45th Street
Seattle, WA 98195

Dr. Ann Bookman
Radcliffe College
The Mary Ingraham Bunting Institute
Ten Garden Street
Cambridge, MA 02138

Dr. David Bowers
Rensis Likert Associates
Wolverine Towers, Suite 401
3001 S. State Street
Ann Arbor, MI 48104

Dr. Jomills Braddock
The Johns Hopkins University
Center for The Social Organization of Schools
3505 Charles Street
Baltimore, MD 21218

Dr. Vincent Carroll
Applied Research Center
Wharton School
University of Pennsylvania
3508 Market Street, Suite 100
Philadelphia, PA 19104

Dr. Terry Connolly
University of Arizona
Department of Psychology, Rm. 312
Tucson, AZ 85721

Dr. James Coyne
Mental Research Institute
555 Middlefield Rd.
Palo Alto, CA 94301

Dr. Richard Daft
Texas A&M University
Department of Management
College Station, TX 77843

Dr. Randy Dunham
University of Wisconsin
Graduate School of Business
Madison, WI 53706

Dr. Cynthia D. Fisher
College of Business Administration
Texas A&M University
College Station, TX 77843

Dr John Frederickson
Boulton, Beranek & Newman, Inc.
10 Moulton Street
Cambridge, MA 02138

Dr. Meg Gerrard
Iowa State University
Psychology Department
Ames, IA 50010

Dr. J. Richard Hackman
School of Organization
    and Management
Box 1A, Yale University
New Haven, CT 06520

Dr. Ronald Heifetz
Harvard University
John F. Kennedy School of Government
79 John F. Kennedy Street
Cambridge, MA 02138

Dr. Wayne Holder
American Humane Associaton
P.O. Box 1266
Denver, CO 80201-1266

Dr. Daniel Ilgen
Department of Psychology
Michigan State University
East Lansing, MI 48824

Dr. Larry James
Georgia Tech Research Corp.
Administration Building
Georgia Institute of Technology
Atlanta, GA 30332

Dr. David Johnson
Professor, Educational Psychology
178 Pillsbury Drive, S.E.
University of Minnesota
Minneapolis, MN 55455

Dr. Robert Rice
State University of New York at Buffalo
Department of Psychology
Buffalo, NY 14226

Dr. Irwin Sarason
University of Washington
Department of Psychology
Seattle, WA 98195

Dr. Edgar Schein
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Benjamin Schneider
University of Maryland
Department of Psychology
College Park, MD 20742

Dr. H. Wallace Sinaiko
Program Director, Manpower Research
    and Advisory Services
Smithsonian Institution
801 N. Pitt Street, Suite 120
Alexandria, VA 22314

Dr. Eliot Smith
Psychology Department
Purdue University
West Lafayette, IN 47907

Dr. Harry C. Triandis
Department of Psychology
University of Illinois
Champaign, IL 61820

Dr. Anne S. Tsui
Duke University
The Fuqua School of Business
Durham, NC 27706

Dr. Andrew H. Van de Ven
University of Minnesota
Office of Research Administration
1919 University Avenue
St. Paul, MN 55104

Dr. Sara Yogev
Northwestern University
Graduate School of Management
2040 Sheridan Road
Evanston, IL 60201

END

FILMED

2-86

DTIC